

基于灵敏度分析的模块化回声状态网络修剪算法

王磊^{1,2} 乔俊飞¹ 杨翠丽¹ 朱心新¹

摘要 针对回声状态网络 (Echo state network, ESN) 的结构设计问题, 提出基于灵敏度分析的模块化回声状态网络修剪算法 (Pruning algorithm for modular echo state network, PMESN). 该网络由相互独立的子储备池模块构成. 首先利用矩阵的奇异值分解 (Singular value decomposition, SVD) 构造子储备池模块的权值矩阵, 并利用分块对角阵原理生成储备池. 然后利用子储备池模块输出和相应的输出层权值向量, 定义学习残差对于子储备池模块的灵敏度以及网络规模适应度. 利用灵敏度大小判断子储备池模块的贡献度, 并根据网络规模适应度确定子储备池模块的个数, 删除灵敏度低的子模块. 在网络的修剪过程中, 不需要缩放权值就可以保证网络的回声状态特性. 实验结果说明, 所提出的算法有效解决了 ESN 的网络结构设计问题, 基本能够确定与样本数据相匹配的网络规模, 具有较好的泛化能力和鲁棒性.

关键词 修剪算法, 模块化回声状态网络, 奇异值分解, 灵敏度分析, 网络规模适应度

引用格式 王磊, 乔俊飞, 杨翠丽, 朱心新. 基于灵敏度分析的模块化回声状态网络修剪算法. 自动化学报, 2019, 45(6): 1136–1145

DOI 10.16383/j.aas.c180288

Pruning Algorithm for Modular Echo State Network Based on Sensitivity Analysis

WANG Lei^{1,2} QIAO Jun-Fei¹ YANG Cui-Li¹ ZHU Xin-Xin¹

Abstract To design the structure of echo state network (ESN), a pruning algorithm for modular echo state network (PMESN) based on sensitivity analysis is proposed in this paper. The reservoir of PMESN is made up of independent sub-reservoir modular networks. The weight matrices of sub-reservoir modular networks are obtained by the singular value decomposition (SVD), and the reservoir is generated by the block diagonal matrix theory. The residual error's sensitivities to the sub-reservoir modular networks are defined by their outputs and weight vectors connecting to the output layer. The significance of the sub-reservoir modular networks is determined by sensitivity. The model scale adaptability is obtained and the sub-reservoir modular networks are sorted by the defined sensitivities. Then, the number of requisite sub-reservoir modular networks is calculated by the model scale adaptability. The redundant sub-reservoir modular networks with smaller sensitivities are deleted. In the pruning process, the echo state property can be guaranteed without posterior scaling of the weights. The simulation results show that the proposed method can design the compact structure of ESN effectively and has better generalization ability and robustness.

Key words Pruning algorithm, modular echo state network, singular value decomposition, sensitivity analysis, model scale adaptability

Citation Wang Lei, Qiao Jun-Fei, Yang Cui-Li, Zhu Xin-Xin. Pruning algorithm for modular echo state network based on sensitivity analysis. *Acta Automatica Sinica*, 2019, 45(6): 1136–1145

回声状态网络 (Echo state network, ESN) 是

一种新型递归神经网络, 它的核心是随机稀疏连接的庞大的储备池^[1]. 如果 ESN 具有回声状态特性 (Echo state property, ESP), 则具有较好的稳定性和动态特性, 因此 ESN 设计的前提是具有 ESP 特性. 在 ESN 的训练过程中, 输入权值和储备池权值一旦生成, 就固定不变, 只需要对输出权值由简单的线性回归进行训练而得到. ESN 这种训练方法的优点可以避免陷入局部极小, 具有较快的学习速度. ESN 已经成功地应用于时间序列预测^[2–3]、非线性控制^[4]、语音识别等领域^[5].

ESN 的储备池是随机设定的, 储备池的这种随机性导致了不清晰的动态特性, 一般难以达到根据具体任务相匹配的最优结构. 储备池规模过大, 虽然可以精确地拟合训练数据, 但会引入大量的冗余特

收稿日期 2018-05-08 录用日期 2018-09-18
Manuscript received May 8, 2018; accepted September 18, 2018
国家自然科学基金 (61533002, 61603012), 北京市教委项目 (KM201710005025), 北京市博士后工作经费资助项目 (2017ZZ-028), 中国博士后科学基金资助

Supported by National Natural Science Foundation of China (61533002, 61603012), Beijing Municipal Education Commission Foundation (KM201710005025), Beijing Postdoctoral Research Foundation (2017ZZ-028) and China Postdoctoral Science Foundation

本文责任编辑 鲁仁全

Recommended by Associate Editor LU Ren-Quan

1. 北京工业大学信息学部 北京 100124 2. 北京信息科技大学高动态导航技术北京市重点实验室 北京 100192

1. Faculty of Information Technology, Beijing University of Technology, Beijing 100124 2. Beijing Key Laboratory of High Dynamic Navigation Technology, Beijing Information Science & Technology University, Beijing 100192

征和无关特征, 容易产生过拟合现象, 影响网络的泛化能力. 反之, 储备池规模过小, 虽然在计算量上有所减少, 然而会影响神经网络的有效性, 易出现欠拟合的现象^[6-7]. 因此如何设计和构造根据具体任务相匹配的储备池结构是 ESN 研究的一个热点和难点.

ESN 储备池设计主要方法有确定型储备池^[2, 8-10]、增长方法^[6]、修剪方法^[7, 11]、正则化方法^[12-13]、进化方法^[14-16]等. 基于确定型储备池拓扑结构的 ESN 极大地改进了传统的 ESN, 克服了传统 ESN 储备池的随机性. Rodan 等探讨了最小复杂度储备池, 并给出了简单环型储备池 (Simple cycle reservoir, SCR), SCR 不但结构简单而且性能不亚于传统的 ESN, 同时 SCR 的动态性能优于传统的 ESN^[8]. 传统回声状态网络具有单个储备池, 储备池神经元存在耦合现象. Xue 等提出了解耦回声状态网络 (Decoupled echo state network, DESN), DESN 的储备池由多个子储备池构成, 但是没有给出确定子储备池个数的方法^[9]. 薄迎春等给出了一种具有小世界特性的回声状态网络 (Small world echo state network, SWESN), SWESN 动力学特性比常规 ESN 更为丰富, 在鲁棒性、抗干扰性等方面均优于传统的 ESN^[10]. 伦淑娟提出了改进的小世界回声状态网络, 改进了储备池神经元节点的连接方式, 有目的地实现了稀疏连接, 减小了储备池随机稀疏连接的盲目性^[2]. 增长方法在初始阶段结构简单, 具有计算量小的优点, 但是合适的停止准则是增长法的难点, 同时在储备池的增长过程中, 要具有 ESP 特性以保证网络的稳定性. Qiao 等给出了增量式模块化回声状态网络 (Growing echo state network, GESN), GESN 的储备池由多个子储备池组成, 子储备池逐渐添加至网络中, 消弱了储备池神经元间的耦合特性. 而且, 在储备池增长过程中, 储备池的稀疏度单调下降, 无需对储备池先验矩阵调整即可满足稳定条件, 也能够减少计算负担^[6]. 修剪方法是从一个较大结构开始, 通过停止准则去掉冗余神经元, 在修剪过程中同样需要具有 ESP 特性. Dutoit 等给出了修剪算法对储备池与输出层的冗余权值进行修剪来提高 ESN 的泛化能力, 但是没有对储备池进行修剪^[7]. Wang 等给出了敏感迭代修剪算法对 SCR 进行修剪 (Sensitive iterated pruning algorithm-simple cycle reservoir, SIPA-SCR), 首先构造一个较大的 SCR, 然后减掉敏感度最低的储备池神经元^[11]. 正则化方法通过正则化参数平衡网络性能与储备池规模, 其难点是正则化参数难以确定. 韩敏等给出了基于 L_1 范数正则化的方法控制网络规模的复杂度^[12]. 结合 L_1 范数和 L_2 范数正则化, Xu 等给出了自适应弹性网算法 (Adaptive elas-

tic echo state network, AEESN) 控制网络规模^[13]. 进化类算法是基于生物进化原理的搜索方法, 具有很好的鲁棒性和全局搜索能力, 适用于神经网络结构的优化和调整. Duan 等在 ESN 的训练阶段利用正交鸽群优化算法 (Orthogonal pigeon-inspired optimization, OPIO) 优化网络参数, 如储备池规模、谱半径、稀疏度等, 实验结果表明 OPIO-ESN 具有较好的网络性能, 且以较快的速度收敛到全局最优^[14]. Zhong 等利用进化算法对双层回声状态网络的参数进行优化, 如储备池规模、谱半径等, 实验结果表明, 改进的 ESN 在精度和稳定性方面具有较好结果^[15]. Wang 等利用二进制粒子群优化方法 (Binary particle swarm optimization, BPSO) 优化 ESN 的输出层连接, 从而改进了传统的 ESN 结构来提高 ESN 的性能与泛化能力^[16].

针对以上问题, 本文提出了基于灵敏度分析的模块化回声状态网络修剪算法, 解决了 ESN 的结构设计问题. 在保证网络稳定性的前提下, 既构建了更加紧凑的网络结构, 又保证了网络的泛化能力.

1 回声状态网络

回声状态网络的储备池内含有成百上千个神经元, 彼此耦合在一起. ESN 的参数较多, 储备池规模、稀疏度、输入权值及储备池权值范围等需要大量尝试甚至是运气. ESN 包含输入层、储备池层、输出层等三层, 没有输出反馈的基本 ESN 的拓扑结构如图 1 所示. 输入权值矩阵和储备池权值矩阵在训练前随机确定, 且一旦生成保持固定不变, 只需要对输出权值进行训练. 没有输出反馈的基本 ESN (Original echo state network, OESN) 方程如下:

$$\mathbf{x}(n+1) = \tanh(W^{\text{in}}\mathbf{u}(n+1) + W\mathbf{x}(n)) \quad (1)$$

$$\mathbf{y}(n+1) = W^{\text{out}}(\mathbf{u}(n+1)^{\text{T}}, \mathbf{x}(n+1)^{\text{T}})^{\text{T}} \quad (2)$$

其中, $\mathbf{u}(n) \in \mathbf{R}^K$, $\mathbf{x}(n) \in \mathbf{R}^N$, $\mathbf{y}(n) \in \mathbf{R}^L$ 分别为 ESN 网络的输入变量, 储备池状态变量和输出变量; $W^{\text{in}} \in \mathbf{R}^{N \times K}$, $W \in \mathbf{R}^{N \times N}$, $W^{\text{out}} \in \mathbf{R}^{L \times (K+N)}$ 分别表示输入权值矩阵、储备池权值矩阵、输出权值矩阵; \tanh 为储备池神经元的激活函数. 为了克服初始瞬态的影响, 假设从 $n_{\min} + 1$ 时刻开始收集内部状态矩阵 $X = [\mathbf{X}(n_{\min} + 1), \mathbf{X}(n_{\min} + 2), \dots, \mathbf{X}(P)]^{\text{T}}$ (P 为训练样本数), 其中 $\mathbf{X}(n) = [\mathbf{u}(n)^{\text{T}}, \mathbf{x}(n)^{\text{T}}]^{\text{T}}$, 对应的期望输出矩阵为 $Z = [\mathbf{z}(n_{\min} + 1), \mathbf{z}(n_{\min} + 2), \dots, \mathbf{z}(P)]^{\text{T}}$ 输出权值矩阵的计算公式由最小二乘法可得

$$W^{\text{out}} = (X^+ Z)^{\text{T}} = ((X^{\text{T}} X)^{-1} X^{\text{T}} Z)^{\text{T}} \quad (3)$$

其中, T 代表矩阵的转置, X^+ 代表 X 的广义逆.

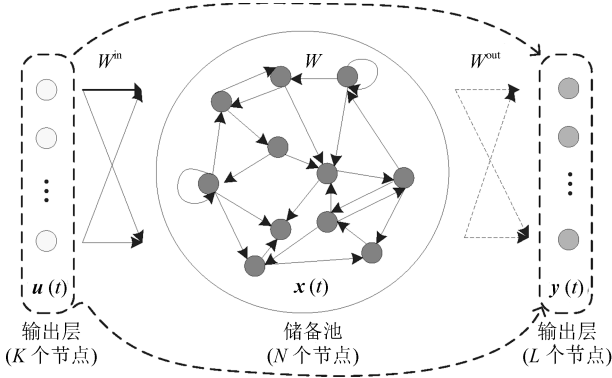


图 1 无输出反馈的基本 ESN 结构

Fig. 1 The basic architecture of the OESN without feedback

2 基于灵敏度分析的修剪型模块化回声状态网络

2.1 具有多个子储备池的模块化回声状态网络

构建具有充分多的子储备池的模块化回声状态网络 (Modular echo state network, MESN), 其储备池由相互独立的子储备池构成, 可以保证储备池部分神经元的解耦, 如图 2 所示。

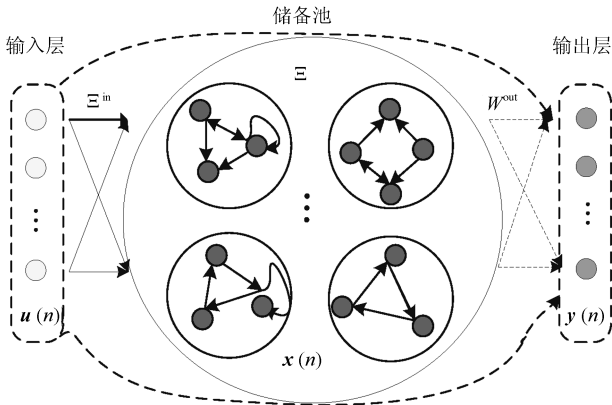


图 2 无输出反馈的具有多个子储备池的模块化 ESN 结构

Fig. 2 The architecture of MESN without feedback

设子储备池个数记为 I 个, MESN 的储备池对应的矩阵为分块对角阵 Ξ , 如式 (4) 所示, W_i ($1 \leq i \leq I$) 为对应第 i 个子储备池的权值矩阵, 对 W_i ($1 \leq i \leq I$) 采用 SVD 分解生成: 首先, 按给定的奇异值分布随机生成对角矩阵 $S_i = \text{diag}\{\sigma_i^1, \sigma_i^2, \dots, \sigma_i^{n_i}\}$ ($1 \leq i \leq I, 0 < \sigma_i^j < 1$), n_i 为第 i 个子储备池的规模, 子储备池的规模可以相同, 也可以不相同, 根据经验一般取为 2~5, 为了处理问题的方便, 假设子储备池规模均相同, 且子储备池全连接; 其次, 同时生成两个随机正交矩阵 $U_i = (u_{jk})_{n_i \times n_i}, V_i = (v_{jk})_{n_i \times n_i}$, 其中 $u_{jk}, v_{jk} \in (-1, 1)$ ($j, k = 1, 2, \dots, n_i$), 计算第 i 个

子储备池的权值矩阵 $W_i = U_i S_i V_i$ ($i = 1, 2, \dots, I$), 由奇异值分解, W_i 与 S_i 有相同的奇异值, 所以 W_i 的所有奇异值均小于 1, 而 Ξ 为分块对角阵, 从而 Ξ 的所有奇异值也均小于 1, 满足回声状态特性的充分性条件. Ξ^{in} 为对应的输入权值矩阵, 如式 (5) 所示, 其中 W_i^{in} ($1 \leq i \leq I$) 为第 i 个子储备池对应的输入权值矩阵. $\Theta(n)$ 为 MESN 的储备池状态变量, 如式 (6) 所示, 其中 $\mathbf{x}_i(n)$ ($1 \leq i \leq I$) 为第 i 个子储备池的状态变量, 其中储备池状态变量更新公式如式 (7) 所示:

$$\Xi = \begin{bmatrix} W_1 & & & \\ & W_2 & & \\ & & \ddots & \\ & & & W_I \end{bmatrix} \quad (4)$$

$$\Xi^{\text{in}} = \begin{bmatrix} W_1^{\text{in}} \\ W_2^{\text{in}} \\ \vdots \\ W_I^{\text{in}} \end{bmatrix} \quad (5)$$

$$\Theta(n) = (\mathbf{x}_1^{\text{T}}(n), \mathbf{x}_2^{\text{T}}(n), \dots, \mathbf{x}_I^{\text{T}}(n))^{\text{T}} \quad (6)$$

$$\Theta(n) = \tanh(\Xi^{\text{in}} \mathbf{u}(n) + \Xi \Theta(n-1)) \quad (7)$$

定义第 i 个子储备池的内部状态矩阵: $H_i = [\mathbf{X}_i(n_{\min} + 1), \mathbf{X}_i(n_{\min} + 2), \dots, \mathbf{X}_i(P)]^{\text{T}}$ ($1 \leq i \leq I$), 其中 $\mathbf{X}_i(n) = [\mathbf{u}^{\text{T}}(n), \mathbf{x}_i^{\text{T}}(n)]^{\text{T}}$ ($n = n_{\min} + 1, \dots, P$), 相应的期望输出矩阵为 $Z = [z(n_{\min} + 1), z(n_{\min} + 2), \dots, z(P)]^{\text{T}}$, P 为训练样本数, 则定义 MESN 的内部状态矩阵 $H = [H_1, H_2, \dots, H_I]$ MESN 相应的网络输出如式 (8) 所示, 输出权值矩阵由最小二乘法计算, 如式 (9) 所示:

$$\mathbf{y}(n) = H(W_*^{\text{out}})^{\text{T}} \quad (8)$$

$$W_*^{\text{out}} = (H^+ Z)^{\text{T}} = ((H^{\text{T}} H)^{-1} H^{\text{T}} Z)^{\text{T}} \quad (9)$$

其中, T 代表矩阵的转置, H^+ 代表 H 的广义逆。

2.2 灵敏度及网络规模适应度的定义

记

$$\mathbf{y}(n) = H_1(W_1^{\text{out}})^{\text{T}}(n) + H_2(W_2^{\text{out}})^{\text{T}}(n) + \dots + H_I(W_I^{\text{out}})^{\text{T}}(n) \quad (10)$$

假设删除第 i 个子储备池模块, 则网络输出变为

$$\mathbf{y}'(n) = H_1(W_1^{\text{out}})^{\text{T}}(n) + \dots + H_{i-1}(W_{i-1}^{\text{out}})^{\text{T}}(n) + H_{i+1}(W_{i+1}^{\text{out}})^{\text{T}}(n) + \dots + H_I(W_I^{\text{out}})^{\text{T}}(n) \quad (11)$$

则残差 $\|\mathbf{y}(n) - \mathbf{y}'(n)\| = \|H_i(W_i^{\text{out}})^T(n)\|$, 因此, 定义学习残差相对于第 i 个子储备池模块的灵敏度为

$$S_i = \frac{1}{P - n_{\min}} \sum_{n=n_{\min}+1}^P \|H_i(W_i^{\text{out}})^T(n)\| \quad (12)$$

S_i 越大表明第 i 个子储备池模块对于学习残差的贡献度越大, 所以由灵敏度可以对子储备池模块进行排序 $S'_1 \geq S'_2 \geq \dots \geq S'_I$. 删除的子储备池模块贡献度越大, 学习残差越大, 所以网络规模适应度 (Model scale adaptability, MSA) 可由灵敏度定义为

$$M_k = \frac{\sum_{i=1}^k S'_i}{I}, 1 \leq k \leq I \quad (13)$$

与给定任务相匹配的网络规模由子储备池模块个数决定, 子模块数可以由网络规模适应度定义为

$$M = \min \{k | M_k \geq \gamma, 1 \leq k \leq I\} \quad (14)$$

其中, γ ($0 \leq \gamma \leq 1$) 为网络规模适应度阈值, M 为保留的子储备池模块的数量, $I - M$ 为冗余的子储备池模块个数, 阈值 γ 可以由凑试的方法确定. 为了防止样本信息因子储备池模块被删除而消失, 被保留的子储备池模块的输入权值需要进一步更新.

2.3 修剪型 MESN

根据灵敏度 S_i 将 I 个子储备池模块排序如下: $S_{i_1}^1 \geq S_{i_2}^2 \geq \dots \geq S_{i_M}^M \geq S_{i_{M+1}}^{M+1} \geq \dots \geq S_{i_I}^I$, 其中, 下标 $1 \leq i_j \leq I$ 为排序前的子储备池模块序号, 上标 $1 \leq i \leq I$ 为排序后的子储备池模块位次序号. 由式 (13) 和 (14) 选择前 M 个子储备池模块予以保留, 后 $I - M$ 个删除. 记 $R = \{i_1, i_2, \dots, i_M\}$ 为保留的子储备池模块标号集, $\bar{R} = \{i_{M+1}, \dots, i_I\}$ 为删除的子储备池模块标号集. 为了保存被删除子储备池模块的样本信息, 同时消除过拟合信息, 采用权值平均横向传播的方法更新已经保留的子储备池模块的输入权值, 更新公式为:

$$W_{i_j, \text{new}}^{\text{in}} = W_{i_j}^{\text{in}} + \frac{1}{I - M} \sum_{k=M+1}^I W_{i_k}^{\text{in}} \quad (15)$$

其中, $1 \leq j \leq M$, $W_{i_j}^{\text{in}}$, $W_{i_j, \text{new}}^{\text{in}}$ 分别为更新前后的第 j 个子储备池模块对应的输入权值矩阵, $W_{i_k}^{\text{in}}$ ($k = M + 1, \dots, I$) 为删除的子储备池模块对应的输入权值矩阵. 式 (15) 意味着在删除了敏感度低的 $I - M$ 个子储备池模块后, 对保留的子储备池模块对应的输入权值进行更新.

基于以上分析, 下面给出修剪型 MESN (Pruning algorithm for modular echo state network, PMESN) 的详细步骤:

步骤 1. 给定 P 个训练样本, 构建一个规模较大的 MESN, 子储备池模块个数记为 I 个, 记内部状态矩阵为 H , 输出层权值矩阵为 W_*^{out} , 设置网络规模适应度阈值 γ ;

步骤 2. 根据 H 和 W_*^{out} , 计算学习残差对第 i 个子储备池模块的灵敏度 S_i , 并对子储备池模块进行排序 $S_{i_1}^1 \geq S_{i_2}^2 \geq \dots \geq S_{i_M}^M \geq S_{i_{M+1}}^{M+1} \geq \dots \geq S_{i_I}^I$, 其中, 下标 $1 \leq i_j \leq I$ 为排序前的子储备池模块序号, 上标 $1 \leq i \leq I$ 为排序后的子储备池模块位次序号;

步骤 3. 根据学习残差灵敏度 S_i , 由式 (13) 计算网络规模适应度 M_k ;

步骤 4. 根据 M_k 和 γ , 由式 (14) 确定子储备池模块个数, 从而选择储备池被删除的子模块进行修剪得到新的 MESN 的储备池权值矩阵 Ξ_{new} ;

步骤 5. 对保留子储备池模块对应的输入权值由式 (15) 进行更新得到 $\Xi_{\text{new}}^{\text{in}}$;

步骤 6. 对修剪后的网络计算新的状态矩阵 H_{new} , 并计算输出权值矩阵 $W_{\text{new}}^{\text{out}}$;

步骤 7. 对训练好的 PMESN 进行测试.

下面分析 PMESN 的计算复杂度. PMESN 的计算复杂度主要分为两个部分:

1) 储备池权值矩阵的构造过程: 初始构造 I 个子储备池的计算复杂度 $O(In^3)$, 计算子储备池模块灵敏度的复杂度 $O(In)$, 网络规模适应度的计算复杂度 $O(I)$, 保留的子储备池模块的数量的计算复杂度 $O(I)$, 其中 n 为子储备池规模, I 为初始子储备池个数. 因此, 该步骤的计算复杂度为 $O(In^3) + O(In) + O(I) + O(I) = O(I(n^3 + n + 2))$.

2) 输出权值矩阵的计算过程: $O((Mn)^3)$, 其中 n 为子储备池规模, M 为保留的子储备池模块的数量.

综合以上两个步骤的分析, PMESN 的计算复杂度为 $O(I(n^3 + n + 2) + (Mn)^3)$. 对比的网络模型 SCR^[8]、DESN^[9]、GESN^[6]、SIPA-SCR^[11] 及 AEESN^[13] 的计算复杂度基本也可以由储备池权值构造和输出权值的计算而产生. PMESN 的计算复杂度高于 OESN、SCR 和 DESN. 第 4 节的仿真实验所需要的训练时间在一定程度上说明了这个问题.

3 稳定性分析

引理 1^[17]. 对于没有输出反馈的 ESN, 储备池激活函数为 \tanh , 如果储备池权值矩阵的最大奇异值 $\sigma_{\max} < 1$, 则 ESN 具有回声状态特性.

定理 1 (稳定性定理). 对于没有输出反馈的

PMESN, 储备池激活函数为 \tanh , 如果子储备池权值矩阵 W_i ($1 \leq i \leq M$) 的奇异值 σ_i^j ($1 \leq j \leq n_i$) 满足 $0 < a \leq \sigma_i^j \leq b < 1$, 则 PMESN 具有回声状态特性.

证明. 根据灵敏度式 (13) 和网络规模适应度式 (14) 对储备池规模进行修剪后的储备池权值矩阵为

$$\begin{aligned} \Xi_{\text{new}} &= \text{diag}\{W_{i_1}, \dots, W_{i_M}\} = \\ &\text{diag}\{U_{i_1} S_{i_1} V_{i_1}, \dots, U_{i_M} S_{i_M} V_{i_M}\} = \\ &U' \text{diag}\{S_{i_1}, \dots, S_{i_M}\} V' = \\ &U' \text{diag}\{\sigma_{i_1}^1, \dots, \sigma_{i_1}^{n_{i_1}}, \dots, \sigma_{i_M}^1, \dots, \sigma_{i_M}^{n_{i_M}}\} V' \end{aligned}$$

其中, $U' = \text{diag}\{U_{i_1}, \dots, U_{i_M}\}$, $V' = \text{diag}\{V_{i_1}, \dots, V_{i_M}\}$ 为正交矩阵. 则 Ξ_{new} 与 $\text{diag}\{S_{i_1}, \dots, S_{i_M}\}$ 有相同的奇异值, 因此 σ_i^j ($i_1 \leq i \leq i_M, 1 \leq j \leq n_i$) 为 Ξ_{new} 的全部奇异值, 又 $0 < a \leq \sigma_i^j \leq b < 1$, 所以 Ξ_{new} 的最大奇异值 $\sigma_{\max} < 1$. 由引理可知, PMESN 具有回声状态特性. \square

4 仿真实验

本节选择 3 个仿真实验对 PMESN 性能进行对比研究, 包括 Lorenz 时间序列预测^[13, 18-19], 非线性系统辨识^[19-20], 污水处理过程出水氨氮浓度预测^[20-21]. 对比的网络模型取 SCR^[8]、DESN^[9]、GESN^[6]、SIPA-SCR^[11] 及 AEESN^[13]. 所有仿真实验均在理想的软硬件环境中进行: 联想台式机 (i7-4790, CPU 3.60 GHz, 8.0 GB RAM), 操作系统 Windows7, Matlab 软件版本 2013b.

为了测试 PMESN 的预测性能和鲁棒性, 分别选取标准均方根误差 (Normalized root-mean-square error, NRMSE) 和实验设计成功率 $R(\theta)$ 作为网络的性能和鲁棒性评价指标^[6], 计算公式分别如下:

$$\text{NRMSE} = \sqrt{\sum_{n=1}^S \frac{(z(n) - y(n))^2}{S\sigma^2}} \quad (16)$$

$$R(\theta) = \frac{\sum_{i=1}^J h(e_i - \theta)}{J} \quad (17)$$

$$h(x) = \begin{cases} 1, & x \leq 0 \\ 0, & x > 0 \end{cases} \quad (18)$$

其中, $z(n)$ 为期望输出, $y(n)$ 为网络输出, S 为测试样本数量, σ^2 为期望输出的方差, J 为蒙特卡罗实验次数, e_i 为第 i 次实验的测试 NRMSE, θ 为误差阈值. $R(\theta)$ 用于估计模型预测误差不大于误差阈值 θ 的概率, $R(\theta)$ 越大说明网络的鲁棒性越好.

4.1 Lorenz 时间序列预测

Lorenz 系统是时间序列预测的经典基准函数^[13, 18-19], 其公式如下

$$\begin{cases} \frac{dx}{dt} = a_1(y - x) \\ \frac{dy}{dt} = -xz + a_2x - y \\ \frac{dz}{dt} = xy - a_3z \end{cases} \quad (19)$$

其中, a_1, a_2 和 a_3 是系统参数, 为了保证 Lorenz 系统具有混沌特性, 这些系统参数的典型取值为 $a_1 = 10, a_2 = 28, a_3 = 8/3$. $x(t), y(t), z(t)$ 是 Lorenz 系统的三维空间向量. 在实验中, 用四阶 Runge-Kutta 方法生成 2500 个样本数据集, 同时, 为了表明实验的可靠性, 数据样本叠加均匀分布在区间 $[-0.001, 0.001]$ 上的随机噪声. 用 Y - 维样本 $y(t)$ 作时间序列预测, 输入为 $[y(t-2), y(t-1), y(t)]^T$, 输出为 $y(t+1)$. 在数据样本中, 前 1500 个样本用于训练, 后 1000 个样本用于测试, 同时为了克服初始瞬态的影响, 丢弃训练样本中的前 500 个. 步长取为 0.01, 初值 $x(0) = 1, y(0) = 1, z(0) = 0$.

为了测试子储备池规模、储备池初始规模以及网络规模适应度阈值对网络性能的影响, 对以上三个参数, 分别选择相应的不同数值的 PMESN 进行测试. 对于以上三个参数的不同取值, 分别进行独立仿真 20 次. 表 1 给出了不同的子储备池规模对网络性能的影响, 由表 1 可以看出, 随着子储备池规模的增大, 测试 NRMSE 的平均值及标准差先减小后增大, 当子储备池规模为 5 时取最小值, 但此时训练时间最长. 基于此, 本文的实验中子储备池规模均取为 5. 表 2 给出了储备池初始规模对网络性能的影响, 储备池初始规模选择区间为 $[100, 600]$, 每间隔 100 取一次. 由表 2 可知, 随着储备池初始规模的增大, 训练时间也逐渐增大, 测试 NRMSE 变化比较平稳, 因此不取过大的储备池初始规模. 表 3 给出了网络规模适应度阈值 γ 对网络性能的影响, γ 从 1 开始逐渐取到 0.1, 每次选取间隔为 0.1. 由表 3 可以看出, 当 γ 逐渐减小时, 储备池最终规模逐渐减小, 测试 NRMSE 在 γ 取 $1 \sim 0.4$ 时变化比较平稳, 当 γ 取 $0.3 \sim 0.1$ 时测试 NRMSE 急剧增大. 这说明前期删除的子储备池模块对网络性能的影响小, 灵敏度能够准确地对子储备池模块进行排序, 可以通过调节 γ 的取值得到结构紧凑的网络模型.

为了进一步验证 PMESN 算法的有效性, 分别与 SCR、DESN、GESN、SIPA-SCR 及 AEESN 做对比研究, 在相同条件下进行 20 次独立仿真. 详细的对比结果如表 4 所示, 由表 4 可以看出, 虽然

PMESN 的训练时间不是最短的, 但是储备池最终规模为 285, 相对其他网络模型具有较紧凑的网络结构, 而且具有较好的预测性能. 图 3 给出了 PMESN 及 OESN 对含噪声的 Lorenz 时间序列预测输出及预测误差, 由图 3 可以看出 PMESN 的性能优于 OESN. 图 4 给出了基于 PMESN 和 OESN 的含噪声的 Lorenz 时间序列的模型设计成功率, 由图 4 可以看出 PMESN 对于含有噪声的 Lorenz 时间序列预测模型设计成功率高于 OESN. 当阈值 θ 为 3×10^{-3} 时, PMESN 的实验设计成功率是 81%, 大约是 OESN 的 2 倍, 这表示 PMESN 具有较高的鲁棒性.

表 1 子储备池规模对 PMESN 性能的影响

Table 1 Influence of sub-reservoir size on PMESN

子储备池规模	训练时间 (s)	测试 NRMSE	
		平均值	标准差
3	25.12	8.92×10^{-3}	8.87×10^{-4}
5	27.23	4.56×10^{-3}	4.82×10^{-4}
10	24.35	5.13×10^{-3}	5.48×10^{-4}
15	23.68	6.35×10^{-3}	6.69×10^{-4}
20	22.19	6.98×10^{-3}	6.95×10^{-4}

表 2 储备池初始规模对 PMESN 性能的影响

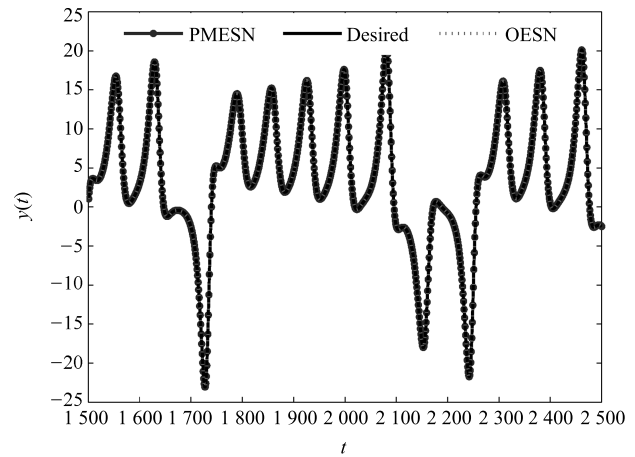
Table 2 Influence of initial reservoir size on PMESN

储备池初始规模	训练时间 (s)	测试 NRMSE	
		平均值	标准差
100	22.15	8.92×10^{-3}	8.78×10^{-4}
200	24.54	4.89×10^{-3}	4.82×10^{-4}
300	26.36	4.72×10^{-3}	4.89×10^{-4}
400	27.26	4.35×10^{-3}	4.48×10^{-4}
500	28.39	3.99×10^{-3}	4.06×10^{-4}
600	32.68	4.82×10^{-3}	5.38×10^{-4}

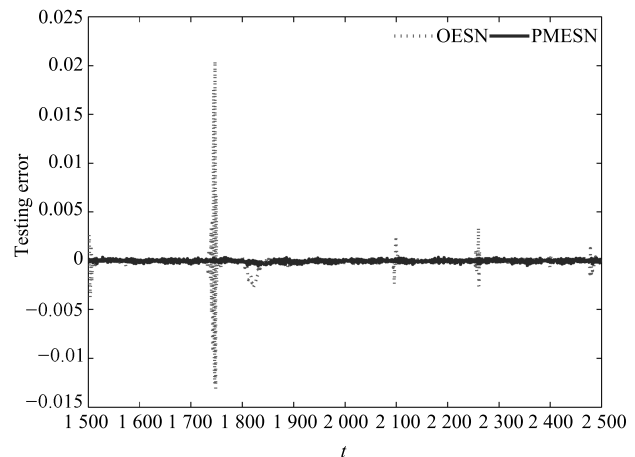
表 3 网络规模适应度阈值对 PMESN 性能的影响

Table 3 Influence of fitness threshold of network size on PMESN

网络规模适应度阈值	储备池初始规模	测试 NRMSE		储备池最终规模
		平均值	标准差	
1	500	4.12×10^{-3}	4.23×10^{-4}	432
0.9	500	4.23×10^{-3}	4.19×10^{-4}	413
0.8	500	4.08×10^{-3}	4.29×10^{-4}	395
0.7	500	4.21×10^{-3}	4.36×10^{-4}	382
0.6	500	4.19×10^{-3}	4.06×10^{-4}	365
0.5	500	4.15×10^{-3}	4.13×10^{-4}	329
0.4	500	4.02×10^{-3}	4.09×10^{-4}	298
0.3	500	9.58×10^{-3}	9.37×10^{-4}	275
0.2	500	5.58×10^{-3}	5.62×10^{-4}	246
0.1	500	8.69×10^{-3}	8.36×10^{-4}	213



(a)



(b)

图 3 基于 PMESN 和 OESN 的含噪声的 Lorenz 时间序列预测结果

Fig. 3 Prediction results based on PMESN and OESN for Lorenz time series with noise

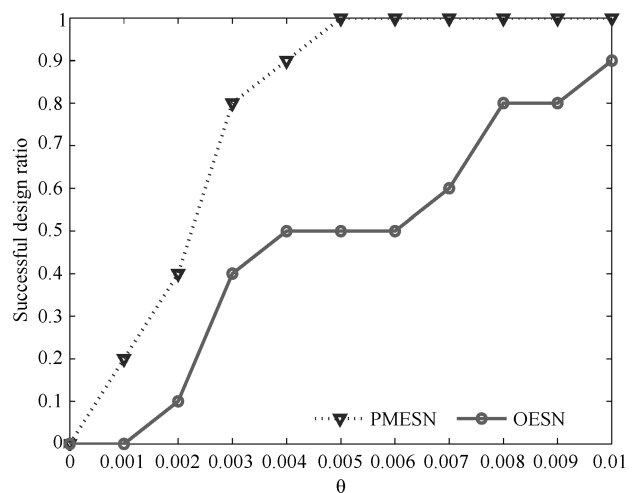


图 4 基于 PMESN 和 OESN 的含噪声的 Lorenz 时间序列的模型设计成功率

Fig. 4 Successful design ratio based on PMESN and OESN for Lorenz time series with noise

表 4 基于不同模型的含噪声的 Lorenz 时间序列预测的参数和仿真结果对比

Table 4 Comparison of some parameters and simulation results of different models for Lorenz time series with noise

网络模型	储备池 初始规模	储备池 最终规模	谱半径	稀疏度	网络规模适 应度阈值	训练时间 (s)	NRMSE	
							平均值	标准差
PMESN	500	285	0.8500	0.0100	0.4	28.85	4.01×10^{-3}	3.64×10^{-4}
OESN ^[1]	500	500	0.8500	0.0500	-	25.32	8.38×10^{-3}	6.38×10^{-4}
SCR ^[8]	500	500	0.8000	0.0020	-	22.15	8.28×10^{-3}	8.16×10^{-4}
DESN ^[9]	500	500	0.8000	0.0238	-	27.35	9.12×10^{-3}	9.43×10^{-4}
GESN ^[6]	50	400	0.9236	0.0200	-	81.35	3.96×10^{-3}	4.15×10^{-4}
SIPA-SCR ^[11]	500	463	0.8500	0.0020	-	41.39	5.65×10^{-3}	5.68×10^{-4}
AEESN ^[13]	500	385	0.8500	0.0500	-	31.39	5.31×10^{-3}	5.06×10^{-4}

“-”表示原文献中无此参数

4.2 非线性系统辨识

系统辨识是现代控制论和信号处理的重要内容, 系统的动态特性被认为表现在变化的输入输出数据中. 而实际系统大都是非线性的, 因此非线性系统辨识成为一个重要而复杂的问题. 非线性系统是系统辨识的标杆问题, 具有大时延、非线性动态的特点, 常用于检验神经网络的建模能力, 其数学模型由式 (20) 给出^[19-20], 其中, 训练阶段的输入 $u(t) = 1.05 \times \sin(t/45)$, 输出的初始值 $y(1) = y(2) = y(3) = y(4) = 0$, 测试阶段的输入如式 (21) 所示.

$$y(t+1) = 0.72y(t) + 0.025y(t-1)u(t-1) + 0.01u^2(t-2) + 0.2u(t-3) \quad (20)$$

$$u(t) = \begin{cases} \sin\left(\frac{\pi t}{25}\right), & 0 < t < 250 \\ 1.0, & 250 \leq t < 500 \\ -1.0, & 500 \leq t < 750 \\ 0.6\left(\frac{\pi t}{10}\right) + 0.1\left(\frac{\pi t}{32}\right) + 0.3\sin\left(\frac{\pi t}{25}\right), & 750 \leq t < 1000 \end{cases} \quad (21)$$

非线性系统通过 $[u(t), y(t)]^T$ 来预测输出 $y(t+1)$. 训练样本数为 1000, 其中训练样本数据集的丢弃点数为 200. 测试样本数为 1000. 同时, 为了验证实验的可靠性, 数据样本叠加均匀分布在区间 $[-0.001, 0.001]$ 上的随机噪声.

在相同条件下, 进行独立仿真 20 次, 详细的参数设置及实验比对结果如表 5 所示, 包括储备池初始规模、储备池最终规模、谱半径、稀疏度、网络规模适应度阈值、训练时间、测试 NRMSE. 由表

5 可以看出, PMESN 具有紧凑的网络结构, 虽然训练时间不是最快的, 但是具有相对较高的测试精度. 图 5 给出了 PMESN 及 OESN 对含噪声的非线

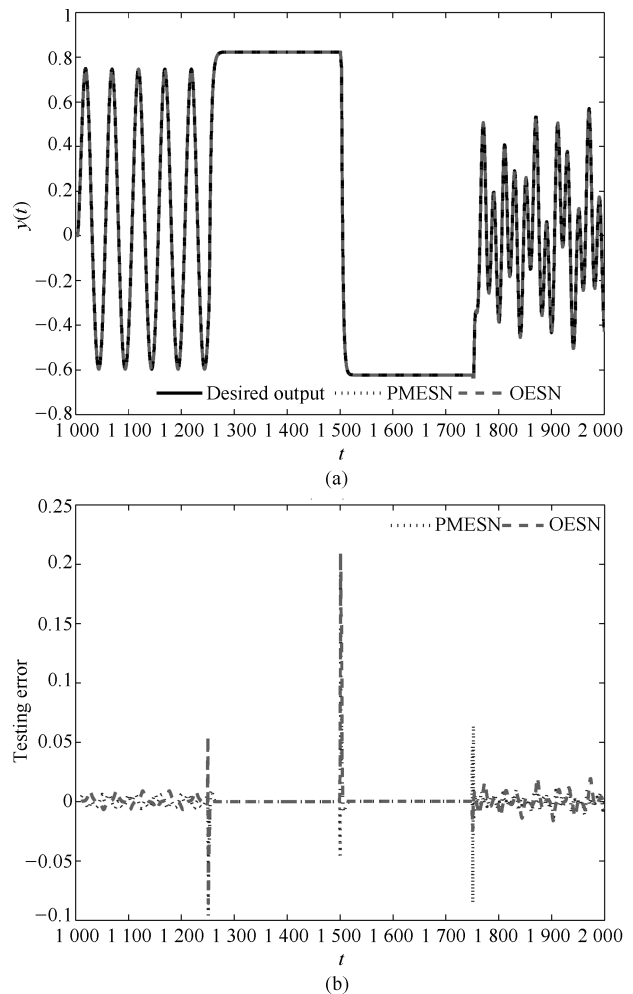


图 5 基于 PMESN 和 OESN 的含噪声的非线性系统辨识预测结果

Fig. 5 Prediction results based on PMESN and OESN for nonlinear system identification with noise

性系统辨识的预测输出及预测误差, 由图 5 可以看出 PMESN 具有较好的网络性能. 图 6 给出了基于 PMESN 和 OESN 的含噪声的非线性系统辨识的模型设计成功率, 当阈值为 0.03 时, PMESN 的实验设计成功率大约是 OESN 的 4 倍, 这说明 PMESN 的鲁棒性较好.

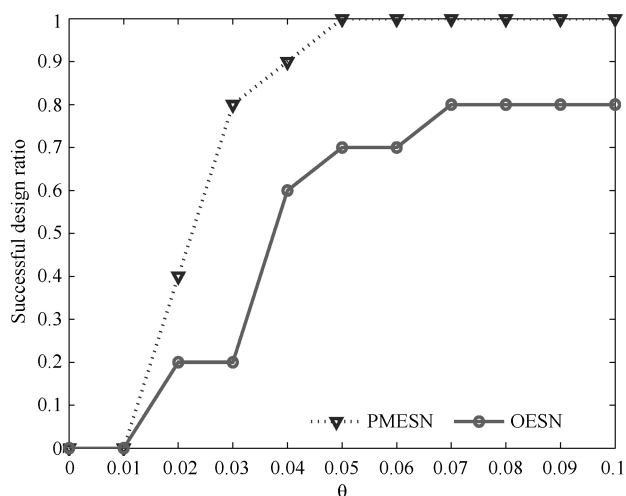


图 6 基于 PMESN 和 OESN 的含噪声的非线性系统辨识的模型设计成功率

Fig. 6 Successful design ratio based on PMESN and OESN for nonlinear system identification with noise

4.3 污水处理过程出水氨氮浓度预测

出水氨氮 (Ammonia nitrogen, $\text{NH}_4\text{-N}$) 浓度是评价污水处理过程出水水质的重要参数. 然而, 出水 $\text{NH}_4\text{-N}$ 浓度难以预测, 现有的检测仪存在成本昂贵, 维护费用高等问题^[20-21]. 在本实验中, PMESN 模型用于预测污水处理过程出水 $\text{NH}_4\text{-N}$ 浓度值, 实验数据来源于北京某污水处理厂 2016 年 6 月 1 日到 2017 年 5 月 31 日期间的数据. 选取易测变量作为模型输入变量, 易测变量有温度、出水 pH 值、出水氧化还原电位 (Oxidation reduction potential, ORP)、溶解氧 (Dissolved oxygen, DO)、总固体悬浮物 (Total suspended solids, TSS). 在剔除异常数据后, 得到 2029 组数据样本并进行归一化处理, 前 1200 组数据样本用于训练, 其中训练集中丢弃点的样本数为 200, 后 829 组数据样本用于测试.

在相同条件下, 进行 20 次独立仿真实验, 详细的参数设置和实验结果如表 6 所示. 由表 6 可以看出 PMESN 储备池最终规模较小, 结构紧凑, 具有较好的泛化能力, 但是训练时间不是最快的. 图 7 给出了出水 $\text{NH}_4\text{-N}$ 浓度的预测结果图, 图 7 说明 PMESN 具有较好的性能. 图 8 给出了基于 PMESN 和 OESN 的模型设计成功率, 当阈值是 0.1 时, OESN 的模型设计成功率是 30%, 而 PMESN 的模型设计成功率是 80%, 说明 PMESN 鲁棒性高.

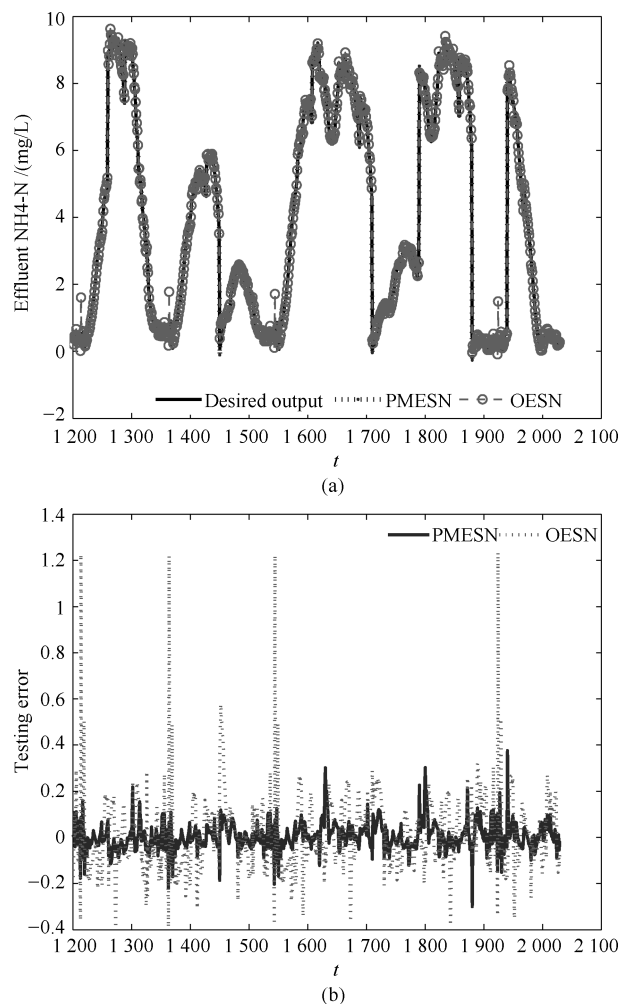


图 7 基于 PMESN 和 OESN 的出水 $\text{NH}_4\text{-N}$ 浓度预测结果
Fig. 7 Prediction results based on PMESN and OESN for effluent $\text{NH}_4\text{-N}$ prediction

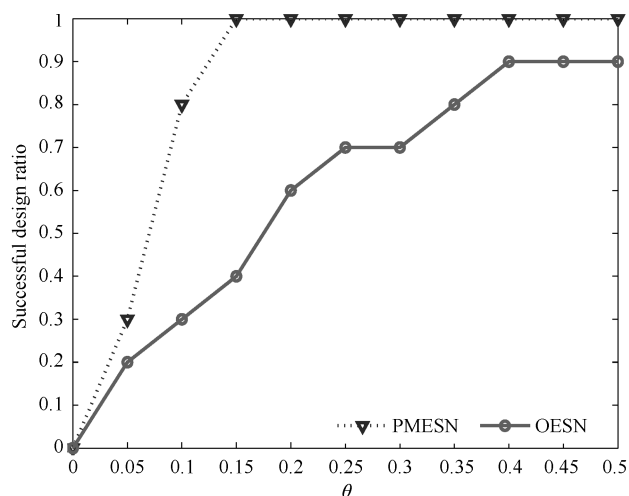


图 8 基于 PMESN 和 OESN 的出水 $\text{NH}_4\text{-N}$ 浓度预测的模型设计成功率

Fig. 8 Successful design ratio based on PMESN and OESN for effluent $\text{NH}_4\text{-N}$ prediction

表 5 基于不同模型的含噪声的非线性系统辨识的参数和仿真结果对比
Table 5 Comparison of some parameters and simulation results of different models for nonlinear system identification with noise

网络模型	储备池	储备池	谱半径	稀疏度	网络规模适	训练时间 (s)	NRMSE	
	初始规模	最终规模			应度阈值		平均值	标准差
PMESN	500	245	0.8500	0.0100	0.5	39.88	0.0359	0.0020
OESN ^[1]	500	500	0.8500	0.0500	—	34.46	0.0723	0.0023
SCR ^[8]	500	500	0.8000	0.0020	—	29.86	0.0692	0.0021
DESN ^[9]	500	500	0.8000	0.0238	—	36.85	0.0812	0.0022
GESN ^[6]	50	400	0.9236	0.0200	—	83.69	0.0436	0.0019
SIPA-SCR ^[11]	500	445	0.8500	0.0020	—	45.66	0.0582	0.0024
AEESN ^[13]	500	376	0.8500	0.0500	—	37.79	0.0519	0.0018

“—”表示原文献中无此参数

表 6 基于不同模型的出水 NH₄-N 浓度预测的参数和仿真结果对比
Table 6 Comparison of some parameters and simulation results of different models for effluent NH₄-N prediction

网络模型	储备池	储备池	谱半径	稀疏度	网络规模适	训练时间 (s)	NRMSE	
	初始规模	最终规模			应度阈值		平均值	标准差
PMESN	500	255	0.8500	0.0100	0.4	38.83	0.2039	0.0198
OESN ^[1]	500	500	0.8500	0.0500	—	32.19	0.3328	0.0232
SCR ^[8]	500	500	0.8000	0.0020	—	29.86	0.2938	0.0286
DESN ^[9]	500	500	0.8000	0.0238	—	35.92	0.3426	0.0312
GESN ^[6]	50	400	0.9236	0.0200	—	91.08	0.2236	0.0022
SIPA-SCR ^[11]	500	458	0.8500	0.0020	—	44.26	0.2935	0.0301
AEESN ^[13]	500	365	0.8500	0.0500	—	39.33	0.2899	0.0268

“—”表示原文献中无此参数

5 结论

本文针对回声状态网络的结构设计问题, 利用奇异值分解首先构造了具有多个相互独立的子储备池的模块化回声状态网络, 并给出了基于灵敏度分析的模块化回声状态网络的修剪算法 PMESN. 然后给出了学习残差对于子储备池模块的灵敏度的计算方法, 并用灵敏度定义了网络规模适应度. PMESN 根据灵敏度大小对子储备池模块进行排序, 同时计算网络规模适应度, 并与网络规模适应度阈值对比, 确定与训练样本相匹配的网络规模. 子储备池模块被删除后, 将被删除子模块的权值平均叠加到其他子模块的权值上, 使得被删除子储备池模块的样本信息得以保留, 从而得到较为紧凑的网络结构. 在网络的修剪过程中, 回声状态特性可以得到保证, 不需要对权值进行缩放. 实验结果说明, 所提出的算法在保证得到结构紧凑的神经网络的同时, 也达到了较高的测试精度, 并且具有较好的稳定性和良好的泛化能力.

References

- 1 Jaeger H, Haas H. Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication. *Science*, 2004, **304**(5667): 78–80
- 2 Lun Shu-Xian, Lin Jian, Yao Xian-Shuang. Time series prediction with an improved echo state network using small world network. *Acta Automatica Sinica*, 2015, **41**(9): 1669–1679
(伦淑娴, 林健, 姚显双. 基于小世界回声状态网的时间序列预测. *自动化学报*, 2015, **41**(9): 1669–1679)
- 3 Qiao J F, Wang L, Yang C L, Gu K. Adaptive Levenberg-Marquardt algorithm based echo state network for chaotic time series prediction. *IEEE Access*, 2018, **6**: 10720–10732
- 4 Bo Y C, Zhang X. Online adaptive dynamic programming based on echo state networks for dissolved oxygen control. *Applied Soft Computing*, 2018, **62**: 830–839
- 5 Skowronski M D, Harris J G. Noise-robust automatic speech recognition using a predictive echo state network. *IEEE Transactions on Audio, Speech and Language Processing*, 2007, **15**(5): 1724–1730
- 6 Qiao J F, Li F J, Han H G, Li W J. Growing echo-state network with multiple subreservoirs. *IEEE Transactions*

- on *Neural Networks and Learning Systems*, 2017, **28**(2): 391–404
- 7 Dutoit X, Schrauwen B, Van Campenhout J, Stroobandt D, Van Brussel H, Nuttin M. Pruning and regularization in reservoir computing. *Neurocomputing*, 2009, **72**(7–9): 1534–1546
- 8 Rodan A, Tino P. Minimum complexity echo state network. *IEEE Transactions on Neural Networks*, 2011, **22**(1): 131–144
- 9 Xue Y B, Yang L, Haykin S. Decoupled echo state networks with lateral inhibition. *Neural Networks*, 2007, **20**(3): 365–376
- 10 Bo Ying-Chun, Qiao Jun-Fei, Zhang Zhao-Zhao. Analysis and design on structure of small world property ESN. *Control and Decision*, 2012, **27**(3): 383–388
(薄迎春, 乔俊飞, 张昭昭. 一种具有 small world 特性的 ESN 结构分析与设计. *控制与决策*, 2012, **27**(3): 383–388)
- 11 Wang H S, Yan X F. Improved simple deterministically constructed cycle reservoir network with sensitive iterative pruning algorithm. *Neurocomputing*, 2014, **145**(18): 353–362
- 12 Han Min, Ren Wei-Jie, Xu Mei-Ling. An improved echo state network via L_1 -norm regularization. *Acta Automatica Sinica*, 2014, **40**(11): 2428–2435
(韩敏, 任伟杰, 许美玲. 一种基于 L_1 范数正则化的回声状态网络. *自动化学报*, 2014, **40**(11): 2428–2435)
- 13 Xu M L, Han M. Adaptive elastic echo state network for multivariate time series prediction. *IEEE Transactions on Cybernetics*, 2016, **46**(10): 2173–2183
- 14 Duan H B, Wang X H. Echo state networks with orthogonal pigeon-inspired optimization for image restoration. *IEEE Transactions on Neural Networks and Learning Systems*, 2016, **27**(11): 2413–2425
- 15 Zhong S S, Xie X L, Lin L, Wang F. Genetic algorithm optimized double-reservoir echo state network for multi-regime time series prediction. *Neurocomputing*, 2017, **238**: 191–204
- 16 Wang H S, Yan X F. Optimizing the echo state network with a binary particle swarm optimization algorithm. *Knowledge-Based Systems*, 2015, **86**: 182–193
- 17 Jaeger H. The “Echo State” Approach to Analysing and Training Recurrent Neural Networks-with an Erratum Note, GMD Report 148, German National Research Center for Information Technology, Bonn, Germany, 2010.
- 18 Lorenz E N. Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, 1963, **20**(2): 130–141
- 19 Lin C J, Chen C H. Identification and prediction using recurrent compensatory neuro-fuzzy systems. *Fuzzy Sets and Systems*, 2005, **150**(2): 307–330
- 20 Han H G, Lu W, Hou Y, Qiao J F. An adaptive-PSO-based self-organizing RBF neural network. *IEEE Transactions on Neural Networks and Learning Systems*, 2018, **29**(1): 104–117
- 21 Haimi H, Mulas M, Corona F, Vahala R. Data-derived soft-sensors for biological wastewater treatment plants: an overview. *Environmental Modelling and Software*, 2013, **47**: 88–107



王磊 北京工业大学信息学部博士研究生. 主要研究方向为神经网络结构设计与优化. 本文通信作者.

E-mail: jade_wanglei@163.com

(WANG Lei Ph.D. candidate at the Faculty of Information Technology, Beijing University of Technology. His research interest covers structure design and optimization of neural networks. Corresponding author of this paper.)



乔俊飞 北京工业大学信息学部教授. 主要研究方向为污水处理过程智能控制, 神经网络结构设计与优化.

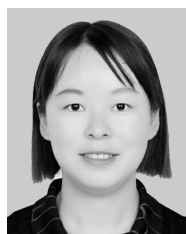
E-mail: junfeq@bjut.edu.cn

(QIAO Jun-Fei Professor at the Faculty of Information Technology, Beijing University of Technology. His research interest covers intelligent control of wastewater treatment process, structure design and optimization of neural networks.)



杨翠丽 北京工业大学信息学部讲师. 主要研究方向为神经网络和智能优化算法. E-mail: clyang5@bjut.edu.cn

(YANG Cui-Li Lecturer at the Faculty of Information Technology, Beijing University of Technology. Her research interest covers neural network and intelligent optimization algorithm.)



朱心新 北京工业大学硕士研究生. 主要研究方向为神经网络结构设计与优化. E-mail: 1205580412@emails.bjut.edu.cn

(ZHU Xin-Xin Master student at the Faculty of Information Technology, Beijing University of Technology. Her research interest covers structure design and optimization of neural networks.)